Identifying and locating-dominating codes in (random) geometric networks

Tobias Müller, Jean-Sébastien Sereni

Identifying and locating-dominating codes in (random) geometric networks*

Tobias Müller[†] Jean-Sébastien Sereni[‡]

Abstract

We model a problem about networks built from wireless devices using identifying and locating-dominating codes in unit disk graphs. It is known that minimising the size of an identifying code is $\mathcal{NP}$-complete even for bipartite graphs. First, we improve this result by showing that the problem remains $\mathcal{NP}$-complete for bipartite planar unit disk graphs. Then, we address the question of the existence of an identifying code for random unit disk graphs. We derive the probability that there exists an identifying code as a function of the radius of the disks and we find that for all interesting ranges of r this probability is bounded away from one. The results obtained are in sharp contrast with those concerning random graphs in the Erdős-Rényi model. Another well-studied class of codes are locating-dominating codes, which are less demanding than identifying codes. A locating-dominating code always exists, but minimising its size is still $\mathcal{NP}$-complete in general. We extend this result to our setting by showing that this question remains $\mathcal{NP}$-complete for arbitrary planar unit disk graphs. Finally, we study the minimum size of such a code in random unit disk graphs, and we prove that with probability tending to one, it is of size $(\frac{n}{r})^{2/3+o(1)}$ if $r \leq \sqrt{2}/2 - \varepsilon$ is chosen such that $nr^2 \to \infty$ and of size $n^{1+o(1)}$ if $nr^2 \ll \ln n$.

1 Introduction

Our results concern two well-studied classes of codes—namely identifying and locating-dominating codes—for (random) unit disk graphs.

Given a graph $G = (V, E)$, we let $\overline{N}(v)$ be the *closed neighbourhood* of the vertex v, that is the set $\{u \in V : uv \in E\} \cup \{v\}$. For any vertex $v \in V$ and subset $C \subseteq V$, the *shadow* of v on C is $\mathrm{Sh}_C(v) := \overline{N}(v) \cap C$. The set C is *covering* if $\mathrm{Sh}_C(v) \neq \emptyset$ for every

*This work was partially supported by the European project IST FET AEOLUS.

[†]School of Mathematics, Tel Aviv University, Ramat Aviv 69978, Israel. E-mail: tobias@post.tau.ac.il.

[‡]CNRS (LIAFA, Université Denis Diderot), Paris, France and Department of Applied Mathematics (KAM), Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic. E-mail: sereni@kam.mff.cuni.cz.

vertex $v \in V$, and C is *separating* if $\mathrm{Sh}_C(u) \neq \mathrm{Sh}_C(v)$ for every pair of distinct vertices $(u, v) \in V^2$. An *identifying code* of G is a set $C \subseteq V$ that is both covering and separating.

Identifying codes are used in several applications. They were first introduced for fault diagnosis of multi-processor systems [11], but they proved to be useful in other areas, in particular location detection in harsh environments (see the work of Ray et al. [16]).

Our work is motivated by the following application. Consider wireless devices scattered around some area that are able to communicate with "users" roaming around the area. If a device decides to handle the communications with a user (this decision is made based on an appropriate protocol) it sends out a signal to nearby devices to indicate that it has become *activated* and has "grabbed" the communications with the user. Because it is quite energy consuming to send a radio signal over a long distance, it makes sense for the activated device to send out the signal only to devices that are within some predefined distance r. A device has only limited power supply (small batteries) and also limited memory. On the other hand, a centralised controller may want to determine which device has grabbed the communications with the user. To do this the controller could of course poll all the devices, but this would be inefficient because long distance radio communications are energy consuming as mentioned above. It would be more efficient if the controller were to poll only a small subset of the devices in such a way that it is able to determine the activated device on the basis of the received information. For similar reasons we want to minimise the amount of information (i.e. number of bits) sent (and stored) by each of the polled devices, since both sending and storing information consumes energy. If each device is associated with a specific region (i.e. it handles communications with all users in the region), as is for instance the case if the grabbing is done according to a nearest distance protocol, then such a system can be used for location detection (i.e. the controller is able to determine which of the regions the user is in). Systems similar to the one described here are used for instance in environmental monitoring, as described by Mainwaring et al. [13].

We model this problem using unit disk graphs. If $V \subseteq \mathbb{R}^2$ and $r > 0$ the *unit disk graph* $G(V, r)$ is a graph with vertex set V and an edge $vw \in E(G(V, r))$ whenever $\| v - w \| < r$. Thus, if we let V correspond to the locations of the devices and r to the range of the activation signal then the objective is to find a (minimum size) identifying code for $G(V, r)$.

Every graph H for which there exists a set $V \subseteq \mathbb{R}^2$ and a positive real r such that $H = G(V, r)$ is a *unit disk graph*, and (V, r) is a *realisation* (or an *embedding*) of H. Up to scaling, the real r can be assumed to be 1. Unit disk graphs have been extensively studied and we refer to the survey by Clark, Colbourn and Johnson [5] for further exposition on this class of graphs.

A graph has an identifying code if and only if the closed neighbourhoods of every two vertices are distinct—if there are two vertices with the same closed neighbourhood, then they cannot be separated, and otherwise the whole set of vertices is an identifying code. Thus, determining whether a given graph admits an identifying code is easy. On the other hand, minimising the size of an identifying code in an arbitrary graph is $\mathcal{NP}$-complete, even when restricted to bipartite graphs [4]. In Subsection 2.1 we strengthen this result by showing that minimising the size of an identifying code in an arbitrary unit disk graph is $\mathcal{NP}$-complete, even when restricted to bipartite planar unit disk graphs.

Section 3 is devoted to random analysis. We consider a sequence $(G_n)_n$ of random unit disk graphs, defined as follows. Points $X_1, X_2, \dots$ are picked uniformly at random from the unit square, and G_n is the graph whose vertex set is $\{X_1, X_2, \dots, X_n\}$, with an edge between two vertices if and only if the corresponding points lie at distance less than r in the plane, where $r = r(n)$ is a sequence of positive distances that may vary with n. These graphs are often also called random geometric graphs and have enjoyed increasing popularity as models for various applications in recent years. We shall determine the (asymptotic) probability that an identifying code exists in terms of r and we shall see that this probability is bounded away from one for all interesting ranges of r. This behaviour is completely different from what happens in the Erdős-Rényi random graph [8]: if p and $1-p$ both are at least $\frac{4\log\log n}{\log n}$, then almost every graph in $\mathcal{G}_{n,p}$ admits an identifying code, and the minimum size $c(n,p)$ of such a code is equivalent to $f(n,p) := \frac{2\log n}{\log(1/(p^2+(1-p)^2))}$, in the sense that for every $\varepsilon > 0$, the probability that $(1-\varepsilon)\cdot f(n,p) < c(n,p) < (1+\varepsilon)\cdot f(n,p)$ tends to 1 when $n \to \infty$.

Our results on the existence of an identifying code in a random unit disk graph indicate that for the relevant applications identifying codes might not work so well in practice. One might therefore want to slightly relax the constraints on the code and perform a similar type of analysis. There is actually a well-studied class of codes that are less demanding, namely locating-dominating codes—see for instance [3, 6, 9]. A locating-dominating code is the same as an identifying code, except that the vertices of the code need not be separated. In fault-diagnosis for instance, this corresponds to the case where some devices can be ensured to be non-faulty. The problem is then to minimise the number of such special devices (which are more expensive). Formally, a *locating-dominating code* of a graph $G = (V, E)$ is a subset C of V such that $\overline{N}(v) \cap C \neq \emptyset$ and $\overline{N}(u) \cap C \neq \overline{N}(v) \cap C$ for every two distinct vertices u, v of $V \setminus C$. Thus, the whole set of vertices is always a locating-dominating code. We prove in section 2.2 that it is $\mathcal{NP}$-complete to minimise the size of a locating-dominating code in an arbitrary planar unit disk graph, and in section 3.2 we establish that, with probability tending to one, the minimum size of a locating-dominating code is $\left(\frac{n}{r}\right)^{2/3+o(1)}$ if $nr^2 \to \infty$, and $n^{1+o(1)}$ if $nr^2 \ll \ln n$. Here and in the rest of the paper, $f(n) \ll g(n)$ means that $\frac{f(n)}{g(n)} \to 0$ when $n \to \infty$.

We end the paper by pointing out some directions for further work on this topic.

2 Complexity

In this section we prove two complexity results about identifying and locating-dominating codes in unit disk graphs.

2.1 Identifying codes in unit disk graphs

Minimising the size of an identifying code is $\mathcal{NP}$-complete for bipartite graphs [4]. We extend this result to arbitrary planar bipartite unit disk graphs.

Theorem 1. *The following problem is $\mathcal{NP}$-complete.*
INSTANCE: *A planar bipartite unit disk graph G along with a realisation of that graph and a positive integer k.*
QUESTION: *Does G admit an identifying code of size at most k?*

The fact that a realisation of the unit disk graph is part of the input is important since determining whether an arbitrary graph is a unit disk graph is $\mathcal{NP}$-complete [2].

We need two lemmas to prove Theorem 1. Given a graph G, a *handle* of G is an induced path of G the vertices of which all have degree 2 in G.

Lemma 2. *Consider a graph G with a handle $P := v_1v_2\dots v_{6k}$ of order $6k$ for a positive integer k. Let x be the neighbour of v_1 in $V(G)\setminus\{v_2\}$. Then, every identifying code C of G contains at least $3k$ vertices of P. Moreover, if C contains exactly $3k$ vertices of P and if $v_{6k}\in C$, then $x\in C$.*

Proof. The proof is by induction on the positive integer k, the result being directly checked if $k=1$. So, suppose that the result is true for an integer $k-1\geq 1$, and let us prove it for k. Let P be a handle as in the statement of the lemma, and C an identifying code of G. The vertices $v_1, v_2, \dots, v_{6(k-1)}$ form a handle P_1 of G, and the vertices $v_{6(k-1)+1}, \dots, v_{6k}$ form a handle P_2 of order 6. By the induction hypothesis, C contains at least $3(k-1)$ vertices of P_1. As the result is true when k is one, C contains at least three vertices of P_2. Therefore, C contains at least $3k$ vertices of P. Moreover, if C contains exactly $3k$ of these vertices, then it contains exactly $3(k-1)$ vertices of P_1 and three vertices of P_2. So, if in addition $v_{6k}\in C$, then $v_{6(k-1)}\in C$, since it is the neighbour of $v_{6(k-1)+1}$ not in P_2. Now, using the induction hypothesis on P_1, we deduce that $x\in C$, as desired. $\square$

The next lemma deals with the property of a particular graph, called a *variable-gadget*.

Definition 3. A *variable-gadget of order m* is the graph $K=(V,E)$ where

- $V := T\cup F\cup R$ with

$$\begin{aligned} T &:= \{t_i : 1\leq i\leq m\}\,, \\ F &:= \{f_i : 1\leq i\leq m\}\,, \text{ and} \\ R &:= \{y_i, z_i : 1\leq i\leq 2m\}\,; \end{aligned}$$

- $E := E_1\cup E_2\cup E_3$ with

$$\begin{aligned} E_1 &:= \{z_{2i-1}t_i : 1\leq i\leq m\}\cup\{z_{2i-1}f_i : 1\leq i\leq m\}\,, \\ E_2 &:= \{z_{2i}t_{i+1} : 1\leq i\leq m\}\cup\{z_{2i}f_i : 1\leq i\leq m\}\,, \text{ and} \\ E_3 &:= \{z_iy_i : 1\leq i\leq 2m\}\,. \end{aligned}$$

Note that for E_2, we set $t_{m+1} := t_1$. See Figure 1 for an example.

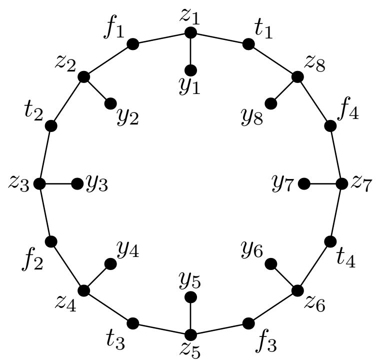


Figure 1: The variable-gadget K of order 4.

Lemma 4. *Consider a graph G containing a variable-gadget K of order m as an induced subgraph. Suppose moreover that only the vertices of $T \cup F$ can have neighbours outside of K. Then, every identifying code C of G contains at least $3m$ vertices of K. Moreover, if C contains exactly $3m$ vertices of K, then either $T \subset C$ and $F \cap C = \emptyset$, or $F \subset C$ and $T \cap C = \emptyset$.*

Proof. Every identifying code C of G must contain, for every $i \in \{1, 2, \ldots, 2m\}$, at least one vertex among y_i, z_i so as to cover y_i. We assert that, for every $i \in \{1, 2, \ldots, 2m\}$, at least one neighbour of z_i different from y_i belongs to C. Otherwise, z_i and y_i are not separated, since the only vertices of C in $\overline{N}(z_i)$ also belong to $\overline{N}(y_i)$. Hence, the number of vertices of C in K is at least $3m$.

Suppose now that C contains a vertex of T and a vertex of F. Then, by the previous remark, observe that there must exist indices i, j, k such that $t_i \in C$, $f_j \in C$ and z_k is adjacent to both t_i and f_j. So we infer that $2 \cdot |C \cap (T \cup F)| \geq 2m + 1$. Therefore C contains at least $m + 1$ vertices in $T \cup F$, which concludes the proof. □

Proof of Theorem 1. Let us outline the proof before going into details. Consider an instance $I = (\varepsilon, X)$ of 3-SAT, where $\varepsilon = (\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m)$ is a set of clauses over the set of variables $X = \{X_1, X_2, \ldots, X_n\}$. We can associate with I the bipartite graph H with vertex-set $\varepsilon \cup X$ and an edge between X_i and C_j whenever $X_i \in C_j$ or $\overline{X_i} \in C_j$. PLANAR 3-SAT is the 3-SAT problem for the class of all instances for which H is planar. It is known that PLANAR 3-SAT is $\mathcal{NP}$-complete [12]. So, consider an instance I of PLANAR 3-SAT.

First we shall compute, in polynomial-time, a particular embedding of H, so-called *box-orthogonal embedding*. Then, we construct from it a planar bipartite unit disk graph G along with a realisation $\mathring{H}$, still in polynomial-time, which has an identifying code of size at most $f(\mathring{H})$ if and only if I can be satisfied. As the function f, to be made precise later, is polynomially computable, this yields the desired result.

A *box-orthogonal embedding* of H is a planar embedding of H such that each edge is represented by alternate horizontal and vertical line segments, and each vertex is represented by a (possibly degenerate) rectangle, called a *box*. All line segments, including those at the perimeter of a box, are assumed to lie on lines of the integer grid—see Figure 2.1. Each planar graph has such an embedding, and it can be computed in polynomial-time [7, 14].

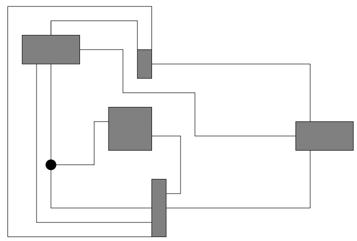


Figure 2: Example of a box-orthogonal embedding.

First, note that every vertex of degree at most 4 can indeed be represented by just a point, and not a (non-degenerated) rectangle: it suffices for this to arrange edges incident to this vertex; see Figure 4(a). Thus, we may assume that all the vertices of ε are represented by points.

We can also ensure that all vertices $X_i \in X$ are represented by rectangles of the same dimensions, which are chosen so that a variable-gadget K_i of order m (and with vertex-set $T_i \cup F_i \cup R_i$) can be embedded on its perimeter. See Figure 3 for the embedding of the gadget on the perimeter of a rectangle. We may assume that all edges are sufficiently long and that there is enough space between the edges for what follows.

The edges around a box B_i are modified as shown in Figure 4(b), so as to ensure that an edge coming from a vertex $\mathcal{C}_j$ reaches a vertex of T_i if $X_i \in \mathcal{C}_j$, and a vertex of F_i if $\overline{X_i} \in \mathcal{C}_j$. This can be done such that for every variable gadget K_i, each vertex of $T_i \cup F_i$ has at most one neighbour outside of K_i. Note that the vertices of R_i have no neighbour outside of K_i.

Now, we compute the length of each edge, and we subdivide each edge by picking points with rational coordinates on the edges in such a way that

- the number of subdivisions is a multiple of 6 for each edge; and
- every two non-consecutive points on an edge are at distance at least $1 + 2\nu$ for some fixed positive rational ν.

(This is possible since we can assume the edges to be sufficiently long.) All these points are added to the vertex set of the graph G we are building. Notice that this step also can be done in polynomial-time.

Last, we add a neighbour o_j to each vertex $\mathcal{C}_j \in \varepsilon$; this does not prevent the graph from being a unit disk graph since the vertices of ε had degree 3. The obtained graph G is a planar unit disk graph, a realisation $\mathring{H}$ being obtained from the planar embedding we built by centring a disk of radius $1/2 + \nu$ at each vertex. It is moreover bipartite, since the following 2-colouring of G is proper. Colour the vertices of ε red and the vertices o_j blue. In each variable-gadget, colour the vertices of $T \cup F \cup \{y_1, y_2, \dots, y_{2m}\}$ blue and the remaining vertices, namely $z_1, z_2, \dots, z_{2m}$, red. Finally, for each path corresponding to an

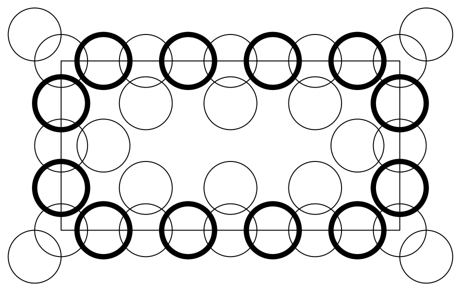
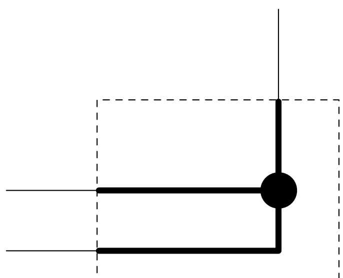
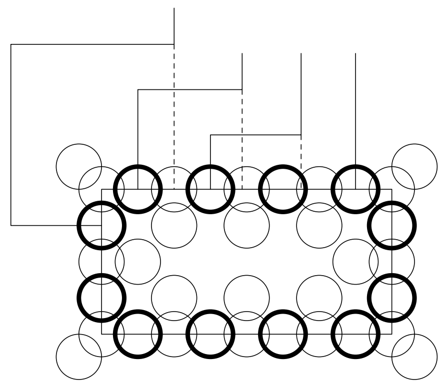


Figure 3: Embedding of a variable-gadget around a box. The bold circles represent the vertices of $T \cup F$.



(a) Reducing vertices of ε to points.



(b) Ensuring that each edge reaches a vertex of T_i or F_i, as desired.

Figure 4: Modifications of the embedding of the edges of H.

edge of H, alternatively colour the vertices red and blue in such a way that the endvertex adjacent to a vertex of a variable-gadget is coloured red, and the other endvertex blue—this is possible because each such path has even order.

We prove now that I can be satisfied if and only if G has a code of size at most $f(\mathring{H})$, defined below. For each clause $\mathcal{C}_j$ three paths, denoted by p_j^ℓ for $\ell \in \{1, 2, 3\}$, join the vertex $\mathcal{C}_j$ to the corresponding literals. Let γ_j^ℓ be the number of internal vertices of the path p_j^ℓ—note that each γ_j^ℓ is of the form $6s$ for some positive integer $s = s(\ell, j)$. Set

$$f(\mathring{H}) := 3nm + m + \frac{1}{2}\sum_{j=1}^{m}\left(\gamma_j^1 + \gamma_j^2 + \gamma_j^3\right) .$$

Suppose first that I can be satisfied. We pick a particular satisfying assignment A and construct an identifying code C of size at most $f(\mathring{H})$. For each variable X_i, the vertices of T_i are added to C if X_i is true, and the vertices of F_i are added to C otherwise. We also add the vertices z_j of R_i. So far C contains $n \times 3m$ vertices. Consider a path $p_j^\ell = xv_1v_2 \ldots v_{6k}\mathcal{C}_j$ where x belongs to a variable-gadget K_i. If the literal to which x

corresponds is true, according to the satisfying assignment A, then $x \in C$ and we add to C the vertices v_{2r} for $r \in \{1, 2, \ldots, 3k\}$. Otherwise, we add to C the vertices v_{2r-1} for $r \in \{1, 2, \ldots, 3k\}$. Last, we add to C the vertices o_j for $j \in \{1, 2, \ldots, m\}$. The obtained code C has size $f(\mathring{H})$. Let us check that C is an identifying code. All the vertices are covered by the definition of C, so it only remains to check that C is separating. To see this, notice that every vertex $\mathcal{C}_j$ has at least one neighbour in C different from o_j, since the clause $\mathcal{C}_j$ is satisfied. Hence $\mathrm{Sh}_C(o_j) = \{o_j\}$ for every $j \in \{1, 2, \ldots, m\}$, while the shadow of $\mathcal{C}_j$ on C consists of o_j and at least one other vertex. The other vertices are surely separated.

Conversely, assume that G has an identifying code C of size at most $f(\mathring{H})$. By Lemmas 2 and 4, the code C contains at least $\gamma_j^\ell/2$ internal vertices of p_j^ℓ, and at least $3m$ vertices in each variable-gadget. Moreover, C must contain at least one vertex among $\mathcal{C}_j, o_j$ so as to cover o_j. Hence, the code C contains exactly that number of vertices in each of the subgraphs mentioned. Thus, by Lemma 4, for each variable-gadget K_i, either $F_i \subset C$ and $T_i \cap C = \emptyset$, or $T_i \subset C$ and $F_i \cap C = \emptyset$. If $T_i \subset C$ then we set the variable X_i to be true, otherwise false. Consider now an arbitrary clause $\mathcal{C}_j$: we infer that at least one neighbour of $\mathcal{C}_j$ different from o_j also belongs to C (otherwise C would not be separating $\mathcal{C}_j$ and o_j). Consider the path $p_j^\ell = xv_1v_2 \ldots v_{6k}\mathcal{C}_j$ to which this vertex belongs: its internal vertices form a handle $v_1v_2 \ldots v_{6k}$ of G. The code C contains exactly $\gamma_j^\ell/2$ vertices of this handle, and $v_{6k} \in C$. Therefore x belongs to C by Lemma 2. By the definition, the vertex x belongs to $T_i \cup F_i$ for some variable-gadget K_i, and hence the corresponding literal is true. Thus the clause $\mathcal{C}_j$ is satisfied. $\square$

2.2 Locating-dominating codes in unit disk graphs

Minimising the size of a locating-dominating code is $\mathcal{NP}$-complete [4]. We extend this result to planar unit disk graphs.

Theorem 5. *The following problem is $\mathcal{NP}$-complete.*
INSTANCE: *A planar unit disk graph G along with a realisation of that graph and a positive integer k.*
QUESTION: *Does G admit a locating-dominating code of size at most k?*

The proof follows the same line as the preceding one, we basically just change the gadgets.

Lemma 6. *Consider a graph G with a handle $P := v_1v_2 \ldots v_{5k}$ of order $5k$ for a positive integer k. Let x be the neighbour of v_1 in $V(G) \setminus \{v_2\}$. Then, every locating-dominating code C of G contains at least $2k$ vertices of P. Moreover, if C contains exactly k of these vertices and if $v_{5k} \in C$, then $x \in C$.*

Proof. The proof is by induction on the positive integer k, so suppose first that $k = 1$. If C contains exactly one vertex of P, it must be v_3, for otherwise one of v_2, v_3, v_4 would not be covered. But then v_2 and v_4 are not separated. So C contains at least two vertices of

P. If C contains exactly two vertices of P, and if $v_5 \in C$, then neither v_1 nor v_4 belongs to C—otherwise v_3 and v_2 would not be covered, respectively. Thus C must contain exactly one of v_2 and v_3. If $x \notin C$, then $v_2 \in C$ since v_1 is covered. Therefore, v_1 and v_3 are not separated, a contradiction. So $x \in C$, as desired.

Suppose now that the result is true for an integer $k-1 \geq 1$: it extends to k in an analogous way as for Lemma 2. $\square$

The next lemma gives a new variable-gadget, designed to deal with locating-dominating codes.

Definition 7. A *variable-gadget of order* m is the graph $L = (V, E)$ where

- $V := T \cup F \cup R$ with

$$\begin{aligned} T &:= \{t_i : 1 \leq i \leq m\}\,, \\ F &:= \{f_i : 1 \leq i \leq m\}\,, \text{ and} \\ R &:= \{x_i, y_i, z_i : 1 \leq i \leq 2m\}\,; \end{aligned}$$

- $E := E_1 \cup E_2 \cup E_3$ with

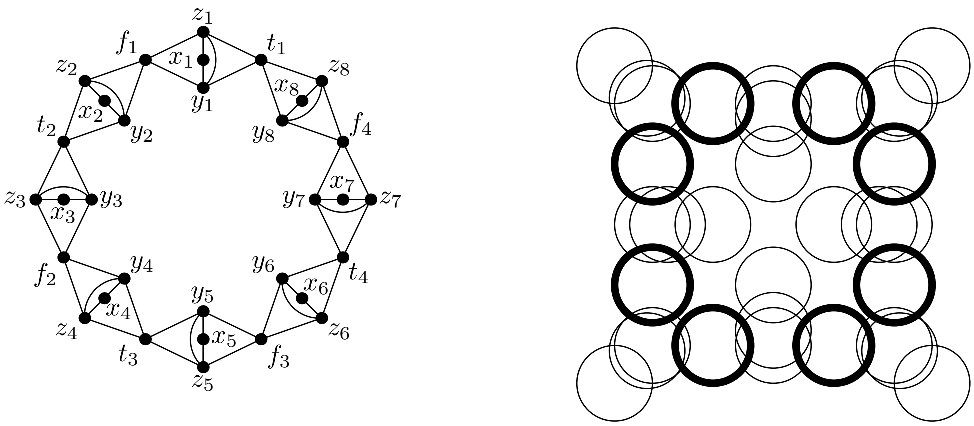
$$\begin{aligned} E_1 &:= \{y_{2i-1}t_i, z_{2i-1}t_i, y_{2i-1}f_i, z_{2i-1}f_i : 1 \leq i \leq m\}\,, \\ E_2 &:= \{y_{2i}t_{i+1}, z_{2i}t_{i+1}, y_{2i}f_i, z_{2i}f_i : 1 \leq i \leq m\}\,, \text{ and} \\ E_3 &:= \{x_iy_i, x_iz_i, z_iy_i : 1 \leq i \leq 2m\}\,. \end{aligned}$$

Note that for E_2, we set $t_{m+1} := t_1$. See Figure 5(a) for an example.

The variable-gadget L is a planar unit disk graph: the disks can be embedded along a box, as illustrated in Figure 5(b).

Lemma 8. *Consider a graph G containing a variable-gadget L as an induced subgraph. Suppose moreover that only the vertices of $T \cup F$ can have neighbours outside of K. Then, every locating-dominating code C of G contains at least $3m$ vertices of K. Moreover, if C contains exactly $3m$ vertices of K, then either $T \subset C$ and $F \cap C = \emptyset$, or $F \subset C$ and $T \cap C = \emptyset$.*

Proof. Every locating-dominating code C of G must contain at least one vertex among y_i, z_i, say y_i, for each $i \in \{1, 2, \ldots, 2m\}$. So C has at least $2m$ vertices of L so far. For each $i \in \{1, 2, \ldots, m\}$, the code must contain at least one vertex among the set $\{x_{2i-1}, z_{2i-1}, t_i, f_i\}$ and at least one among the set $\{x_{2i}, z_{2i}, t_{i+1}, f_i\}$ (where again we set $t_{m+1} := t_1$) in order to separate x_{2i-1} and z_{2i-1} and, respectively, x_{2i} and z_{2i}. Thus, the code C contains at least m additional vertices of L, as desired. Moreover, it contains exactly m additional vertices of L if and only if it contains either all of T (and thus none of F) or all of F. $\square$



(a) (b)

Figure 5: The variable-gadget L of order 4 along with a realisation where the bold circles represent the vertices of $T \cup F$.

Proof of Theorem 5. The proof is analogous to the one of Theorem 1. We keep the same construction, except that we replace the variable gadgets K_i by the variable gadgets L_i, and we make the number of internal vertices of each path joining a variable gadget to a clause vertex a multiple of 5 instead of 6. Recall that for each clause $\mathcal{C}_j$ three paths, denoted by p_j^ℓ for $\ell \in \{1, 2, 3\}$, join the vertex $\mathcal{C}_j$ to the corresponding literals. Let γ_j^ℓ be the number of internal vertices of the path p_j^ℓ: each γ_j^ℓ is thus of the form $5s$ for some positive integer $s = s(\ell, j)$. Recall that for each $j \in \{1, 2, \ldots, m\}$, there is a vertex o_j whose only neighbour is the clause vertex $\mathcal{C}_j$. We add a new vertex o'_j whose only neighbour is o_j. Set

$$f(\mathring{H}) := 3nm + m + \frac{2}{5}\sum_{j=1}^{m}\left(\gamma_j^1 + \gamma_j^2 + \gamma_j^3\right) .$$

Suppose first that the considered instance of PLANAR 3-SAT can be satisfied. We pick a satisfying assignment A and we construct a locating-dominating code C of size at most $f(\mathring{H})$. For each variable X_i, the vertices of T_i are added to C if X_i is true, and the vertices of F_i are added to C otherwise. We also add the vertices y_j of R_i. So far we have $n \times 3m$ vertices in C. Consider a path $p_j^\ell = xv_1v_2 \ldots v_{5k}\mathcal{C}_j$, where x belongs to the variable-gadget L_i. If the literal to which x corresponds is true, according to the satisfying assignment A, then $x \in C$ and we add to C the vertices v_{5r-2} and v_{5r} for $r \in \{1, 2, \ldots, k\}$. Otherwise, we add to C the vertices v_{5r-3} and v_{5r-1} for $r \in \{1, 2, \ldots, k\}$. Last, we add to C the vertices o_j for $j \in \{1, 2, \ldots, m\}$. The obtained code C has size $f(\mathring{H})$. Let us check now that C is a locating-dominating code. That C is covering directly follows from its definition. Now, notice that every vertex $\mathcal{C}_j$ has at least one neighbour in C different from o_j, since the clause $\mathcal{C}_j$ is satisfied. Thus $\mathcal{C}_j$ is separated from o'_j, since $\mathrm{Sh}_C(o'_j) = \{o_j\}$. The other vertices that are not in C are surely separated.

Conversely, assume that G has a locating-dominating code C of size at most $f(\mathring{H})$. By Lemmas 6 and 8, the code C contains at least $2\gamma_j^\ell/5$ internal vertices of p_j^ℓ, and at least

$3m$ vertices in each variable-gadget. Moreover, C must contain at least one vertex among $\mathcal{C}_j, o_j, o'_j$ so as to cover o_j. Hence, the code C contains exactly that number of vertices in each of the subgraphs mentioned. Notice that the only vertex of C among $\mathcal{C}_j, o_j, o'_j$ cannot then be $\mathcal{C}_j$, for otherwise o'_j would not be covered. Also, by Lemma 8, for each variable-gadget L_i, either $F_i \subset C$ and $T_i \cap C = \emptyset$, or $T_i \subset C$ and $F_i \cap C = \emptyset$. If $T_i \subset C$ then we set X_i to be true, otherwise false. Consider now an arbitrary clause $\mathcal{C}_j$: we infer that at least one neighbour of $\mathcal{C}_j$ different from o_j also belongs to C; otherwise C would not be separating $\mathcal{C}_j$ and o_j, or would not cover $\mathcal{C}_j$. Consider the path $p_j^\ell = xv_1v_2 \dots v_{5k}\mathcal{C}_j$ to which this vertex belongs: its internal vertices form a handle $v_1v_2 \dots v_{5k}$ of G. The code C contains exactly $2\gamma_j^\ell/5$ vertices of this handle, and $v_{5k} \in C$. Therefore x belongs to C by Lemma 6. By the definition, the vertex x belongs to $T_i \cup F_i$ for some variable-gadget L_i, and hence the corresponding literal is true. Thus the clause $\mathcal{C}_j$ is satisfied. □

3 Random unit disk graphs

In this section, we consider the random unit disk graph G_n described in the introduction. Furthermore, to simplify the computations we make the *toroïdal convention*, i.e. we identify opposite sides of $[0,1]^2$, making it into a torus. Here distances are measured in the obvious way (formally, we may redefine $\| x \| := \sqrt{\min(x_1, 1-x_1)^2 + \min(x_1, 1-x_2)^2}$ for $x = (x_1, x_2) \in [0,1)^2$).

3.1 Identifying codes

We shall prove the following theorem.

Theorem 9. *The following hold for G_n under the assumptions stated.*

$$\lim_{n\to\infty} \mathbb{P}(G_n \text{ has an ID-code}) = \begin{cases} 1 & \text{if } nr^2 \ll n^{-1}\,, \\ \exp[-\frac{\pi\lambda}{2}] & \text{if } nr^2 \sim \lambda n^{-1}, \text{ for some } \lambda > 0\,, \\ 0 & \text{if } n^{-1} \ll nr^2 \ll n\,, \\ \exp[-\mu(r)] & \text{if } r \text{ is fixed in } (0, \frac{1}{2}\sqrt{2})\,, \\ 0 & \text{if } r \geq \frac{1}{2}\sqrt{2}\,. \end{cases}$$

where for $r \in \left(0, \frac{1}{2}\right)$, we set $\mu(r) := \frac{\pi}{16r^2}$, and for $r \in \left(\frac{1}{2}, \frac{1}{2}\sqrt{2}\right)$, we set

$$\begin{aligned} \mu(r) :=& \frac{1}{4r^2\sin^2\left(\frac{\beta}{2}\right)}\left[\frac{\cos\left(\frac{\beta}{2}\right)}{\cos\left(\frac{\beta}{2}\right)+\sin\left(\frac{\beta}{2}\right)} - \frac{1}{2}\right] \\ &+ \frac{1}{4r^2\sin(\beta)}\left[\frac{2\left(\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)\right)\tan\left(\frac{\beta}{4}\right)}{\left(1-\cos\left(\frac{\beta}{2}\right)+\sin\left(\frac{\beta}{2}\right)\right)\tan^2\left(\frac{\beta}{4}\right)+1+\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)}\right. \\ &\left.+\frac{2}{\sqrt{\sin(\beta)}}\arctan\left(\sqrt{\frac{\left(1-\cos\left(\frac{\beta}{2}\right)+\sin\left(\frac{\beta}{2}\right)\right)}{\left(1+\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)\right)}}\tan\left(\frac{\beta}{4}\right)\right)\right], \end{aligned}$$

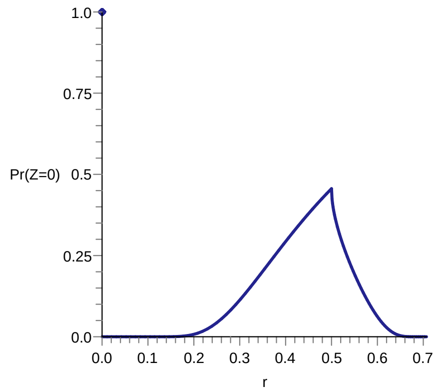


Figure 6: The (asymptotic) probability that an identifying code exists as a function of r, for r fixed.

with $\beta = \beta(r) := \frac{\pi}{2} - 2\arccos\left(\frac{1}{2r}\right)$.

The expression for $\mu(r)$ when $r > \frac{1}{2}$ given in Theorem 9 can be rewritten in terms of r using the relations

$$\cos\left(\frac{\beta}{2}\right) = \frac{1}{2}\sqrt{2}\left(\frac{1}{2r} + \sqrt{1 - \frac{1}{4r^2}}\right), \quad \sin\left(\frac{\beta}{2}\right) = \frac{1}{2}\sqrt{2}\left(\frac{1}{2r} - \sqrt{1 - \frac{1}{4r^2}}\right),$$

together with $\tan\left(\frac{\beta}{4}\right) = \sqrt{\frac{1-\cos(\beta/2)}{1+\cos(\beta/2)}}$. Unfortunately, it does not appear possible to obtain a substantially simpler expression than the one given in Theorem 9.

It should be mentioned that (unbeknownst to them) Agarwal and Spencer also studied the probability that an ID-code exists under a different setting [1]. More precisely, Theorem 9 extends a result of theirs which corresponds to the case where r is fixed in $\left(0, \frac{1}{10}\right)$.

3.1.1 The proof of Theorem 9.

We say that two distinct vertices X_i and X_j of G_n form a *bad pair* if $\overline{N}(X_i) = \overline{N}(X_j)$. Let Z be the number of bad pairs in G_n. Thus

$$\mathbb{P}(G_n \text{ has an ID-code}) = \mathbb{P}(Z = 0)\,. \tag{1}$$

The very first assertion of Theorem 9 is rather trivial.

Lemma 10. *If* $nr^2 \ll n^{-1}$ *then* $\mathbb{P}(Z = 0) = 1 + o(1)$.

Proof. Notice that if G_n contains no edges at all then the whole set of vertices is an identifying code. Let $Y := |E(G_n)|$ be the number of edges of G_n. We see that

$$\mathbb{P}(Y > 0) \leq \mathbb{E}Y \leq \binom{n}{2}\pi r^2 = o(1)\,.$$

So $|E(G_n)| = 0$ with probability $1 + o(1)$. $\square$

Our next aim is to prove the theorem for r in the range $nr^2 \sim \frac{\lambda}{n}$. For this purpose we use the following theorem by Penrose [15].

Theorem 11 (Penrose)**.** *Let $k \in \mathbb{N}$ and suppose $(nr^2)^k \sim \lambda n^{-1}$ for some $\lambda > 0$. Then*

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(\Delta(G_n) = k-1) &= 1 - \lim_{n\to\infty} \mathbb{P}(\Delta(G_n) = k) \\ &= \exp\left[-\tfrac{\lambda}{(k+1)!} \int\limits_{(\mathbb{R}^2)^k} h_k(\{0, x_1, \ldots, x_k\})\,\mathrm{d}x_1 \ldots \mathrm{d}x_k\right], \end{aligned}$$

where $h_k(A)$ equals 1 if $\Delta(G(A,1)) \geq k$ and 0 otherwise.

Originally Theorem 11 was phrased for arbitrary dimension d and every absolutely continuous probability distribution on $\mathbb{R}^d$, but we have taken $d = 2$ and the uniform distribution on $[0,1]^2$ here. We should also mention that we are considering random points on the unit torus, and the theorem was phrased only for probability distributions on $\mathbb{R}^d$. One can however show that if Δ_T is the maximum degree of the random geometric graph on n points drawn uniformly at random from the unit square with the toroïdal convention and Δ_S is the maximum degree of the random geometric graph on the same n points but without the toroïdal assumption, then $\mathbb{P}(\Delta_T \neq \Delta_S) \leq 4r$. This shows that the toroïdal convention does not affect the conclusion of Theorem 11, since there $r \to 0$. To see that $\mathbb{P}(\Delta_T \neq \Delta_S) \leq 4r$, let $d_T(X_i)$ and $d_S(X_i)$ be the degrees of X_i in the random geometric graph with and without the toroïdal convention, respectively. Note that $d_T(X_i) \geq d_S(X_i)$ for all i. Moreover, if $d_T(X_i) \neq d_S(X_i)$ then X_i must be within r of the boundary. Hence, if $\Delta_T \neq \Delta_S$ then all points that satisfy $d_T(X_i) = \Delta_T$ must lie within r of the boundary. Let us pick U uniformly at random from $\{X_i : d_T(X_i) = \Delta_T\}$. By symmetry considerations U has the uniform distribution on $[0,1]^2$. We see that

$$\mathbb{P}(\Delta_T \neq \Delta_S) \leq \mathbb{P}\left(U \text{ is within } r \text{ of the boundary of } [0,1]^2\right) \leq 4r\,,$$

as required. A similar argument will also show that the toroïdal convention does not affect the conclusion of Theorem 9 in the cases where $r \to 0$.

Theorem 11 allows us to give a short proof of the following statement.

Lemma 12. *If $nr^2 \sim \frac{\lambda}{n}$ then $\mathbb{P}(Z = 0) \to e^{-\frac{\pi\lambda}{2}}$.*

Proof. We first assert that whp. the order of each component is at most 2. To see this, let W be the collection of all 3-tuples $(X_{i_1}, X_{i_2}, X_{i_3}) \in \{X_1, \ldots, X_n\}^3$ with i_1, i_2, i_3 distinct and $\| X_{i_1} - X_{i_2} \|, \| X_{i_1} - X_{i_3} \| < r$. Then

$$\mathbb{E}W = n(n-1)(n-2)(\pi r^2)^2 = \mathcal{O}\left(n^{-1}\right)\,,$$

which proves the assertion. Thus, whp. G_n is comprised of isolated vertices and isolated edges. It follows that

$$\mathbb{P}(Z=0) = \mathbb{P}(\Delta(G_n)=0) + o(1)\,.$$

Hence applying Theorem 11 with $k=1$ yields the result, since $\int\limits_{\mathbb{R}^2} h_1(\{0,x\})\,\mathrm{d}x = \pi$. $\square$

For convenience, we split the case when $nr^2 \gg n^{-1}$ and $r=o(1)$ into two subcases, which require/allow different proof techniques.

Lemma 13. *If $n^{-1} \ll nr^2 \ll \ln n$ then $\mathbb{P}(Z=0)=o(1)$.*

Proof. Let Y be the number of isolated edges. As $2\pi r^2 \leq \mathrm{vol}(B(x_1,r)\cup B(x_2,r)) \leq \pi r^2$ for all x_1, x_2, we obtain

$$\binom{n}{2}\pi r^2(1-2\pi r^2)^{n-2} \leq \mathbb{E}Y \leq \binom{n}{2}\pi r^2(1-\pi r^2)^{n-2}\,. \tag{2}$$

Since $\ln(1-x) = -x+\mathcal{O}(x^2)$ and $r=o(1)$, we deduce that $\binom{n}{2}\pi r^2 e^{-(2\pi+o(1))nr^2} \leq \mathbb{E}Y \leq \binom{n}{2}\pi r^2 e^{-(\pi+o(1))nr^2}$. So in particular $\mathbb{E}Y \geq \frac{\pi}{3}n^2r^2e^{-3\pi nr^2}$ (for n sufficiently large). The function $xe^{-3\pi x}$ is increasing for $x\leq(3\pi)^{-1}$ and decreasing for $x\geq(3\pi)^{-1}$. Thus, for any $C\in\mathbb{R}$ we have $C/n \leq nr^2 \leq \ln n/100$ for n sufficiently large. Hence, for n sufficiently large

$$\mathbb{E}Y \geq \frac{\pi}{3}n\cdot\min\left(\frac{C}{n}\cdot e^{-3\pi\frac{C}{n}}, \frac{\ln n}{100}\cdot e^{-3\pi\frac{\ln n}{100}}\right) \geq C\,.$$

This shows that $\mathbb{E}Y\to\infty$.

We now assert that $\mathrm{Var}(Y) = o\left((\mathbb{E}Y)^2\right)$. The conclusion then follows by Chebyschev's inequality, since

$$\mathbb{P}(Z=0) \leq \mathbb{P}(Y=0) \leq \mathbb{P}(|Y-\mathbb{E}Y|\geq\mathbb{E}Y) \leq \frac{\mathrm{Var}(Y)}{(\mathbb{E}Y)^2} = o(1)\,.$$

Thus, it only remains to prove the assertion.

Let us set $\mathcal{P} := \binom{[n]}{2}$ and for $P=\{i,j\}\in\mathcal{P}$ let $I(P)$ be the indicator variable of the event that $\{X_i,X_j\}$ spans an isolated edge. We infer that

$$\begin{aligned} Y^2 &= \sum_{\substack{P_1,P_2\in\mathcal{P}\\|P_1\cap P_2|=2}} I(P_1)I(P_2) + \sum_{\substack{P_1,P_2\in\mathcal{P}\\|P_1\cap P_2|=1}} I(P_1)I(P_2) + \sum_{\substack{P_1,P_2\in\mathcal{P}\\|P_1\cap P_2|=0}} I(P_1)I(P_2)\\ &= \sum_{P\in\mathcal{P}} I(P_1) + \sum_{\substack{P_1,P_2\in\mathcal{P}\\|P_1\cap P_2|=0}} I(P_1)I(P_2)\,. \end{aligned}$$

Here we have used that $I\in\{0,1\}$ and two isolated edges cannot meet in a single vertex. Let $x_1,x_2,x_3,x_4\in[0,1]^2$ be such that $\|x_1-x_2\|< r$, $\|x_3-x_4\|< r$ and $\|x_1-x_3\|> 4r$. If

$P_1 = \{X_1, X_2\}, P_2 = \{X_3, X_4\}$ then, setting $V_r(x,y) := \mathrm{vol}(B(x,r) \cup B(y,r))$,

$$\begin{aligned}\mathbb{E}(I(P_1)I(P_2)|X_1 = x_1, \ldots, X_4 = x_4) &= (1 - V_r(x_1,x_2) - V_r(x_3,x_4))^{n-4} \\ &\leq (1 - V_r(x_1,x_2))^{n-4}(1 - V_r(x_3,x_4))^{n-4} \\ &= \mathbb{P}(I(P_1)|X_1 = x_1, \ldots, X_4 = x_4) \\ &\quad \times \mathbb{P}(I(P_2)|X_1 = x_1, \ldots, X_4 = x_4)\,.\end{aligned}$$

It follows that $\mathbb{E}I(P_1)I(P_2) \leq (\mathbb{E}I(P_1))^2 + 16\pi^3 r^6$ if $P_1 \cap P_2 = \emptyset$. Thus,

$$\mathbb{E}Y^2 = \mathbb{E}Y + \binom{n}{2}\binom{n-2}{2}\mathbb{E}I(P_1)I(P_2) \leq \mathbb{E}Y + (\mathbb{E}Y)^2 + O(n^4 r^6)\,. \tag{3}$$

Consequently, $\mathrm{Var}(Y) \leq \mathbb{E}Y + \mathcal{O}(n^4r^6) = o((\mathbb{E}Y)^2)$ as asserted, since $\mathbb{E}Y \to \infty$ and $n^4r^6/(\mathbb{E}Y)^2 = \mathcal{O}\left(r^2 e^{6\pi n r^2}\right) = \mathcal{O}\left(n^{-1+o(1)}\right)$ using that $r = o\left(\sqrt{\frac{\ln n}{n}}\right)$. $\square$

In the sequel, we set $(m)_k := \frac{m!}{(m-k)!} = m(m-1)\ldots(m-k+1)$.

Lemma 14. *Suppose that $nr^2 \to \infty$ yet $r < \frac{1}{2}\sqrt{2} - \varepsilon$ for some $\varepsilon > 0$. Then*

$$\mathbb{E}(Z)_k = (1 + o(1))\mu(r)^k$$

as $n \to \infty$ for any fixed k.

Once we have established this lemma, Theorem 9 follows easily for the sequences r that have not yet been considered.

Corollary 15. *The following statements hold as $n \to \infty$.*

(i) If $nr^2 \to \infty$ but $r = o(1)$, then $\mathbb{P}(Z = 0) \to 0$.

(ii) If r is fixed in $\left(0, \frac{1}{2}\sqrt{2}\right)$, then $\mathbb{P}(Z = 0) \to e^{-\mu(r)}$.

Proof. Part (i). Notice that in this case $\mathbb{E}Z = (1+o(1))\mu(r) \to \infty$ as $n \to \infty$. Furthermore, since $\mathbb{E}(Z)_2$ and $(\mathbb{E}Z)^2$ are both $(1 + o(1))\mu(r)^2$ by Lemma 14, we have

$$\mathrm{Var}(Z) = \mathbb{E}(Z)_2 - (\mathbb{E}Z)^2 + \mathbb{E}Z = o(\mu(r)^2) = o((\mathbb{E}Z)^2\,.$$

So, again using Chebyschev's Inequality, we indeed find that

$$\mathbb{P}(Z = 0) \leq \frac{\mathrm{Var}(Z)}{(\mathbb{E}Z)^2} = o(1)\,.$$

Part (ii). The Inclusion-Exclusion Principle gives that for every integer l

$$\sum_{k=0}^{2l+1}(-1)^k\frac{\mathbb{E}(Z)_k}{k!} \leq \mathbb{P}(Z = 0) \leq \sum_{k=0}^{2l}(-1)^k\frac{\mathbb{E}(Z)_k}{k!}\,. \tag{4}$$

By Lemma 14, the lower bound in (4) equals $(1 + o(1))\sum_{k=0}^{2l+1}\frac{(-\mu(r))^k}{k!}$ and the upper bound equals $(1 + o(1))\sum_{k=0}^{2l+1}\frac{(-\mu(r))^k}{k!}$. The statement now follows by letting $k \to \infty$, as $\sum_{k=0}^{\infty}\frac{(-\mu(r))^k}{k!} = e^{-\mu(r)}$. $\square$

Before we can give the proof of Lemma 14, we need to do some more ground work. Let $D_r(x,y)$ be the area of the symmetric difference $B(x,r)\Delta B(y,r)$. This difference only depends on $\| y-x \|$, and the angle between $y-x$ and the line $\{(a,a) : a \in \mathbb{R}\}$. By a slight abuse of notation, we also write $D_r(u,\alpha)$ for $D_r(x,y)$ if $u = \| y-x \|$ and α is the angle between $y-x$ and the line $\{(a,a) : a \in \mathbb{R}\}$. For small r $D_r(u,\alpha)$ depends only on u, but for larger r the fact that things take place on $[0,1]^2$ with opposite edges identified makes α relevant. The computations below make use of the following lemma.

Lemma 16. *If $0 < r < \frac{1}{2}$ then $D_r(u,\alpha) = 4ur + \mathcal{O}(u^2)$.*
If $\frac{1}{2} \le r < \frac{1}{2}\sqrt{2}$ then

$$D_r(u,\alpha) = \begin{cases} 4ur\sin\left(\frac{\beta}{2}\right)(\cos(\alpha)-\sin(\alpha)) + o(u) & \text{if } -\frac{\pi}{4} \le \alpha < -\frac{\beta}{2}\,, \\ 4ur\left(1-\left(\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)\right)\cos(\alpha)\right) + o(u) & \text{if } -\frac{\beta}{2} \le \alpha \le 0\,. \end{cases}$$

Here $\beta = \beta(r) := \frac{\pi}{2} - 2\arccos(\frac{1}{2r})$ as before. Furthermore, the error terms $O(u^2)$ and $o(u)$, respectively, can be bounded uniformly in α.

We postpone the proof of Lemma 16 until the end of this section.

Proof of Lemma 14. The proof is by induction on k.

Base case, $k=1$. Let us write

$$F_r(\alpha) := \begin{cases} 1 & \text{if } r \le \frac{1}{2}\,, \\ \sin\left(\frac{\beta}{2}\right)(\cos(\alpha)-\sin(\alpha)) & \text{if } r > \frac{1}{2} \text{ and } -\frac{\pi}{4} < \alpha < -\frac{\beta}{2}\,, \\ 1-\left(\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)\right)\cos(\alpha) & \text{if } r > \frac{1}{2} \text{ and } -\frac{\beta}{2} \le \alpha \le 0\,. \end{cases}$$

Here it is understood that $F_r(\alpha) = F_r(\alpha + \frac{\pi}{4})$. Therefore, by Lemma 16, $D_r(u,\alpha) = 4urF_r(\alpha) + o(u)$ (for $-\frac{\pi}{4} < \alpha \le 0$). Now, observe that $F_r(\alpha) > c$ for some $c = c(\varepsilon)$ uniformly in all r considered, since $r < \frac{1}{2}\sqrt{2} - \varepsilon$. By Lemma 16, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that, if $u < \delta r$, then $(4-\varepsilon)urF(\alpha) < D_r(u,\alpha) < (4+\varepsilon)urF_r(\alpha)$. Now notice that $D_r(u,\alpha) = \Omega(r^2)$ for $u > \delta r$.

$$\begin{aligned} \mathbb{P}(\overline{N}(X_1) = \overline{N}(X_2)) &= \int_0^{2\pi}\int_0^{\delta r}(1-D_r(u,\alpha))^{n-2}u\,\mathrm{d}u\,\mathrm{d}\alpha \\ &\quad + \int_0^{2\pi}\int_{\delta r}^{r}(1-D_r(u,\alpha))^{n-2}u\,\mathrm{d}u\,\mathrm{d}\alpha \\ &\le \int_0^{2\pi}\int_0^{\delta r}(1-D_r(u,\alpha))^{n-2}u\,\mathrm{d}u\,\mathrm{d}\alpha \\ &\quad + \pi r^2 e^{-\Omega(nr^2)}\,. \end{aligned} \tag{5}$$

We shall see later on that the last term on the last line is negligibly small compared to the first term on the last line, but first we must compute the first term in the last line. Observe that

$$\begin{aligned} \int_0^{2\pi}\int_0^{\delta r}(1-D_r(u,\alpha))^{n-2}u\,\mathrm{d}u\,\mathrm{d}\alpha &\le \int_0^{2\pi}\int_0^{\delta r} e^{-(n-2)(4-\varepsilon)urF(\alpha)}u\,\mathrm{d}u\,\mathrm{d}\alpha \\ &= \int_0^{2\pi}\int_0^{\delta(4-\varepsilon)r^2(n-2)F(\alpha)} \frac{e^{-v}v\,\mathrm{d}v\,\mathrm{d}\alpha}{((4-\varepsilon)(n-2)rF(\alpha))^2} \\ &\le \frac{1}{(4-\varepsilon)^2(n-2)^2r^2}\int_0^{2\pi}\frac{1}{F(\alpha)^2}\,\mathrm{d}\alpha \\ &\le \frac{1+o(1)}{(4-\varepsilon)^2n^2r^2}\int_0^{2\pi}\frac{1}{F(\alpha)^2}\,\mathrm{d}\alpha\,, \end{aligned} \tag{6}$$

where we have used the substitution $v = (n-2)(4-\varepsilon)urF(\alpha)$ in the second line and in the third line we have used that $\int_0^\infty te^{-t}\,\mathrm{d}t = 1$.
Provided δ was chosen sufficiently small, $u \le \delta r$ implies that

$$(1-(4+\varepsilon)urF(\alpha)) \ge e^{-(4+2\varepsilon)urF(\alpha)}$$

uniformly in α, as $\ln(1-(4+\varepsilon)urF(\alpha)) = -(4+\varepsilon)urF(\alpha)+\mathcal{O}(u^2r^2)$ and $c(\varepsilon) \le F(\alpha) \le 1$. Computations analogous to (6) thus give that

$$\int_0^{2\pi}\int_0^{\delta r}(1-D_r(u,\alpha))^{n-2}u\,\mathrm{d}u\,\mathrm{d}\alpha \ge \frac{1+o(1)}{(4+\varepsilon)^2n^2r^2}\int_0^{2\pi}\frac{1}{F(\alpha)^2}\,\mathrm{d}\alpha\,. \tag{7}$$

Now, notice that

$$\frac{\pi r^2}{r^{-2}n^{-2}}e^{-\Omega(nr^2)} = \pi\left(nr^2\right)^2e^{-\Omega(nr^2)} = o(1), \tag{8}$$

as nr^2 tends to infinity. So indeed, the second term on the last line of (5) is negligibly small compared to the first. Thus, by symmetry considerations we obtain

$$\mathbb{P}(\overline{N}(X_1) = \overline{N}(X_2)) = \frac{1}{2n^2r^2}(1+o(1))\int_{-\frac{\pi}{4}}^{0}F(\alpha)^{-2}\,\mathrm{d}\alpha\,.$$

In other words, $\mathbb{E}Z = \binom{n}{2}\mathbb{P}(\overline{N}(X_1) = \overline{N}(X_2)) = \frac{1}{4r^2}(1+o(1))\int_{-\frac{\pi}{4}}^{0}F(\alpha)^{-2}\,\mathrm{d}\alpha$.

It remains to determine $\int_{-\frac{\pi}{4}}^{0}F(\alpha)^{-2}\,\mathrm{d}\alpha$. For $r \le \frac{1}{2}$ it equals $\frac{\pi}{4}$, which gives the result. Thus, assume now that $r \in \left(\frac{1}{2},\frac{\sqrt{2}}{2}\right)$. Notice that

$$\begin{aligned}
\int_{-\frac{\pi}{4}}^{-\frac{\beta}{2}}F(\alpha)^{-2}\,\mathrm{d}\alpha &= \sin\left(\tfrac{\beta}{2}\right)^{-2}\int_{-\frac{\pi}{4}}^{-\frac{\beta}{2}}(\cos(\alpha)-\sin(\alpha))^{-2}\,\mathrm{d}\alpha\\
&= \sin\left(\tfrac{\beta}{2}\right)^{-2}\left[\tfrac{\cos(\alpha)}{\cos(\alpha)-\sin(\alpha)}\right]_{-\frac{\pi}{4}}^{-\frac{\beta}{2}}\\
&= \sin\left(\tfrac{\beta}{2}\right)^{-2}\left(\tfrac{\cos\left(\frac{\beta}{2}\right)}{\cos\left(\frac{\beta}{2}\right)+\sin\left(\frac{\beta}{2}\right)}-\tfrac{1}{2}\right).
\end{aligned}$$

For convenience, we let $c := \left(\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)\right)^{-1} = \left(2-\frac{1}{2r^2}\right)^{-\frac{1}{2}}$. We can now write

$$\begin{aligned}
\int_{-\frac{\beta}{2}}^{0}(F(\alpha))^{-2}\,\mathrm{d}\alpha &= \int_{-\frac{\beta}{2}}^{0}\left(1-\left(\cos\left(\tfrac{\beta}{2}\right)-\sin\left(\tfrac{\beta}{2}\right)\right)\cos(\alpha)\right)^{-2}\,\mathrm{d}\alpha\\
&= c^2\int_{-\frac{\beta}{2}}^{0}(c-\cos(\alpha))^{-2}\,\mathrm{d}\alpha\\
&= c^2\left[\tfrac{2\tan\left(\frac{\alpha}{2}\right)}{(c^2-1)((c+1)\tan^2\left(\frac{\alpha}{2}\right)+c-1)}+\tfrac{2c\arctan\left(\sqrt{\frac{c+1}{c-1}}\tan\left(\frac{\alpha}{2}\right)\right)}{(c^2-1)^{\frac{3}{2}}}\right]_{-\frac{\beta}{2}}^{0}\\
&= \tfrac{c^2}{4(c^2-1)}\left(\tfrac{2\tan\left(\frac{\beta}{4}\right)}{(c+1)\tan^2\left(\frac{\beta}{4}\right)+c-1}+\tfrac{2c\arctan\left(\sqrt{\frac{c+1}{c-1}}\tan\left(\frac{\beta}{4}\right)\right)}{\sqrt{c^2-1}}\right)\\
&= \tfrac{1}{\sin(\beta)}\left(\tfrac{2\left(\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)\right)\tan\left(\frac{\beta}{4}\right)}{\left(1-\cos\left(\frac{\beta}{2}\right)+\sin\left(\frac{\beta}{2}\right)\right)\tan^2\left(\frac{\beta}{4}\right)+1+\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)}\right.\\
&\quad\left.+\tfrac{2}{\sqrt{\sin(\beta)}}\arctan\left(\sqrt{\tfrac{\left(1-\cos\left(\frac{\beta}{2}\right)+\sin\left(\frac{\beta}{2}\right)\right)}{\left(1+\cos\left(\frac{\beta}{2}\right)-\sin\left(\frac{\beta}{2}\right)\right)}}\tan\left(\tfrac{\beta}{4}\right)\right)\right).
\end{aligned}$$

The statement follows for $k = 1$.

Induction step: $k > 1$. Let us set $\rho = \rho(n) := \min(r, \frac{\ln^2 n}{nr})$ and let Y be the number of bad pairs X_i, X_j with $\| X_i - X_j \| < \rho$. Notice that $\rho = r$ if $r \le \frac{\ln n}{\sqrt{n}}$ and $r > \rho$ otherwise. We first assert that

$$\mathbb{E}(Z)_k = \mathbb{E}(Y)_k + o(1) \tag{9}$$

for all fixed k. Note that (9) is true if $\rho = r$, so that we can restrict ourselves to the case where $r \ge \rho$. Since the factorial moments can be written as linear combinations of the ordinary moments, it suffices to observe that for any (fixed) k

$$\mathbb{E}Z^k - \mathbb{E}Y^k = \mathcal{O}\left(n^{2k} e^{-\Omega(\ln^2 n)}\right) = o(1)\,.$$

Here we have used that if $\| x - y \| \ge \rho$ then $\mathrm{vol}(B(x,r)\Delta B(y,r)) = \Omega(\rho r)$ by lemma 16. So (9) holds, and therefore it suffices to consider Y instead of Z in the rest of the proof.

Let us assume that the statement of the lemma holds for $k-1$, with $k \ge 2$. Let $\mathcal{P} = \binom{\{X_1,\ldots,X_n\}}{2}$ be the set of all pairs of nodes. For every $P \in \mathcal{P}$, we let $J(P)$ be the event that P is a bad pair and the points of the pair are at distance $< \rho$. Then,

$$\mathbb{E}(Y)_k = \sum_{P_1,\ldots,P_k \in \mathcal{P} \text{ distinct}} \mathbb{P}(J(P_1),\ldots,J(P_k))\,. \tag{10}$$

First notice that the contribution by terms with $P_i \cap P_j \ne \emptyset$ for some $i \ne j$ is small. Indeed,

$$\begin{aligned}
\sum_{\substack{P_1,\ldots,P_k \in \mathcal{P} \text{ distinct} \\ P_i \cap P_j \ne \emptyset \text{ for some } i,j \in \{1,\ldots,k\}}} \mathbb{P}(J(P_1),\ldots,J(P_k)) &\le k^2 \sum_{\substack{P_1,\ldots,P_k \in \mathcal{P} \text{ distinct} \\ |P_{k-1} \cap P_k| = 1}} \mathbb{P}(J(P_1),\ldots,J(P_k)) \\
&\le 2nk^2 \sum_{P_1,\ldots,P_{k-1} \in \mathcal{P} \text{ distinct}} \mathbb{P}(J(P_1),\ldots,J(P_{k-1}))\pi\rho^2 \\
&= \mathcal{O}\left(\mu(r)^{k-1} n\rho^2\right)\,.
\end{aligned}$$

Here we have used the induction hypothesis in the last line. Next notice that if $r \le \frac{\ln n}{\sqrt{n}}$ then $n\rho^2 = nr^2 \le \ln^2 n \ll \sqrt{n}/\ln n \le r^{-1}$; and, if $r > \frac{\ln n}{\sqrt{n}}$ then $n\rho^2 = \ln^4 n/(nr^2) \ll r^{-1}$, as $nr \ge \sqrt{n}\ln n \gg \ln^4 n$. Since $\mu(r) = \Omega(r^{-1})$ this shows that $n\rho^2 = o(\mu(r))$ and hence also $\mu(r)^{k-1} n\rho^2 = o(\mu(r)^k)$. So we see that

$$\begin{aligned}
\mathbb{E}(Y)_k &= \binom{n}{2}\cdots\binom{n-2(k-1)}{2}\mathbb{P}\left(\bigcap_{j=1}^k J(\{X_{2j-1}, X_{2j}\})\right) + o(\mu(r)^k) \\
&= (1+o(1))n^{2k}2^{-k}\mathbb{P}\left(\bigcap_{j=1}^k J(\{X_{2j-1}, X_{2j}\})\right) + o(\mu(r)^k)\,.
\end{aligned}$$

Next, let ρ' equal $4r$ if $r \le \ln^{-10} n$ and $n^{-\frac{49}{100}}$ otherwise. For $i \in \{1,2,\ldots,k\}$, let us set

$$A(i) := \left\{ \bigcap_{j=1}^{i} J(\{X_{2j-1}, X_{2j}\}) \text{ and } \| X_{2i} - X_{2j} \| > \rho' \text{ for all } 1 \le i < j \le i \right\}\,.$$

Notice that

$$\begin{gathered} n^{2k}2^{-k}\mathbb{P}\left(\bigcap_{j=1}^{k} J(\{X_{2j-1},X_{2j}\}) \text{ but not } A(k)\right) \\ \leq \\ n^{2k}2^{-k}\binom{k}{2}\mathbb{P}\left(\bigcap_{j=1}^{k-1} J(\{X_{2j-1},X_{2j}\})\right)\pi^2(\rho')^2\rho^2 \\ = \\ \binom{k}{2}(1+o(1))\mu(r)^{k-1}\pi^2 n^2(\rho')^2\rho^2 \\ = \\ o(\mu(r)^k)\,. \end{gathered}$$

Here we have used that if $r \leq \frac{\ln n}{\sqrt{n}}$ then $n^2(\rho')^2\rho^2 \leq 16n^2r^4 \leq 16\ln^4 n \ll \sqrt{n}/\ln n \leq r^{-1}$; if $\frac{\ln n}{\sqrt{n}} < r < \ln^{-10} n$ then $n^2(\rho')^2\rho^2 = 16n^2r^2\rho^2 = 16\ln^4 n \ll \ln^{10} n \leq r^{-1}$; and if $r \geq \ln^{-10} n$ then $n^2(\rho')^2\rho^2 = n^{-\frac{98}{100}}r^{-2}\ln^4 n \leq n^{-\frac{98}{100}}\ln^{24} n = o(1) \ll r^{-1}$. Consequently,

$$\mathbb{E}(Y)_k = (1+o(1))n^{2k}2^{-k}\mathbb{P}(A(k)) + o(\mu(r)^k)\,, \tag{11}$$

For $x_1,\ldots,x_{2k} \in [0,1]^2$, let us set

$$D_r(x_1,\ldots,x_{2k}) := \mathrm{vol}\left(\bigcup_{i=1}^{k}(B(x_{2i-1},r)\Delta B(x_{2i},r))\right)\,.$$

Our next aim is to show that if $\| x_{2i-1}-x_{2i} \| < \rho$ for every $i \in \{1,2,\ldots,k\}$ and $\| x_{2i}-x_{2j} \| > \rho'$ for all $i,j \in \{1,2,\ldots,k\}$ with $i<j$, then

$$D_r(x_1,\ldots,x_{2k}) = \sum_{i=1}^{k} D_r(x_{2i-1},x_{2i}) + o(n^{-1})\,, \tag{12}$$

where the error term is uniform over all $x_1,\ldots,x_{2k}$ considered. First notice that this is trivial for $r \leq \ln^{-10} n$, as then $\rho' = 4r$ so that $D_r(x_1,\ldots,x_{2k}) = \sum_{i=1}^{k} D_r(x_{2i-1},x_{2i})$ as the sets $B(x_{2i-1},r)\Delta B(x_{2i},r)$ are disjoint for $i=1,\ldots,k$ in this case.

To see this that (12) also holds when $r \geq \ln^{-10} n$, consider $C := (B(x_1,r)\Delta B(x_2,r)) \cap (B(x_3,r)\Delta B(x_4,r))$ under the assumptions that $\| x_1 - x_2 \| < \rho$, $\| x_3 - x_4 \| < \rho$ and $l := \| x_2 - x_4 \| \geq \rho'$. Then C is contained in the intersection of the two annuli $A_2 := \{y : r-\rho < \| y - x_2 \| < r+\rho\}$ and $A_4 := \{y : r-\rho < \| y - x_4 \| < r+\rho\}$, see figure 7 below. We use the bound $\mathrm{vol}(A_2 \cap A_4) \leq 2\frac{\alpha_2}{2\pi}\mathrm{vol}(A_2)$ with α_1,α_2 as shown in figure 7. First notice that

$$\mathrm{vol}(A_2) = \pi((r+\rho)^2 - (r-\rho)^2) = \mathcal{O}(r\rho)\,.$$

Now, the angles α_1,α_2 satisfy

$$\begin{aligned} \cos(\alpha_1) &= \tfrac{l_2}{r+\rho}\,, \quad \cos(\alpha_1+\alpha_2) = \tfrac{l_1}{r-\rho}\,, \\ \sin(\alpha_1) &= \tfrac{h}{r+\rho}\,, \quad \sin(\alpha_1+\alpha_2) = \tfrac{h}{r-\rho}\,. \end{aligned}$$

where h,l_1,l_2 are as in figure 7. In particular,

$$l = l_1 + l_2, \quad h^2 = (r-\rho)^2 - l_1^2 = (r+\rho)^2 - l_2^2\,. \tag{13}$$

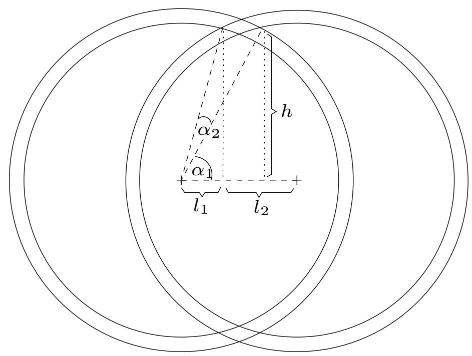


Figure 7: Bounding $\mathrm{vol}(A_2 \cap A_4)$.

Notice that l_1 may be negative (this can happen for small l), but this does not pose any limitation for our computations. First, suppose that $l_2 \geq \frac{1}{2}r$. The Taylor expansion of $\sin(x)$ around $x = \alpha_1$ gives

$$\begin{aligned}\frac{2\rho h}{(r-\rho)(r+\rho)} = \frac{h}{r-\rho} - \frac{h}{r+\rho} &= \sin(\alpha_1+\alpha_2) - \sin(\alpha_1)\\ &= \cos(\alpha_1)\alpha_2 + \mathcal{O}(\alpha_2^2)\\ &= \frac{l_2}{r+\rho}\alpha_2 + \mathcal{O}(\alpha_2^2)\,.\end{aligned}$$

Since $\frac{1}{2}r \leq l_2 \leq r+\rho$ and $h \leq r$, we see that is this case we must have $\alpha_2 = \mathcal{O}(\rho)$.

Now assume that $l_2 \leq \frac{1}{2}r$. The Taylor expansion of $\cos(x)$ around $x = \alpha_1$ gives that

$$\begin{aligned}\frac{r(l_1-l_2)+\rho(l_1+l_2)}{(r-\rho)(r+\rho)} = \frac{l_1}{r-\rho} - \frac{l_2}{r+\rho} &= \cos(\alpha_1+\alpha_2) - \cos(\alpha_1)\\ &= -\sin(\alpha_1)\alpha_2 + \mathcal{O}(\alpha_2^2)\\ &= -\frac{h}{r+\rho}\alpha_2 + \mathcal{O}(\alpha_2^2)\,.\end{aligned}$$

Observe that $h = \sqrt{(r+\rho)^2 - l_2^2} \geq r\frac{1}{2}\sqrt{3}$ and that $l(l_2-l_1) = l_2^2 - l_1^2 = (r+\rho)^2 - (r-\rho)^2 = 4\rho r$, by the relations (13). Hence, $l_2 - l_1 = \mathcal{O}\left(\frac{r\rho}{l}\right) = \mathcal{O}\left(\frac{r\rho}{\rho'}\right)$, since $l \geq \rho'$. Thus, this time we obtain $\alpha_2 = \mathcal{O}\left(\frac{\rho}{\rho'}\right)$.

Consequently, $\mathrm{vol}(A_2 \cap A_4) \leq \frac{\alpha_2}{\pi}\,\mathrm{vol}(A_2) = \mathcal{O}\left(r\rho\cdot\frac{\rho}{\rho'}\right) = \mathcal{O}\left(\frac{\ln^{14} n}{n^{151/100}}\right) = o\left(n^{-1}\right)$. This proves (12).

Now let $x_2, x_4, \ldots, x_{2k} \in [0,1]^2$ be such that $\| x_{2j} - x_{2i} \| > \rho'$ for all $i, j \in \{1, 2, \ldots, k\}$

with $i < j$. We now see that we can write

$$\begin{array}{c}
\mathbb{P}(A(k)|X_2 = x_2, \dots, X_{2k} = x_{2k}) \\
= \\
\int\limits_{B(x_2,\rho)\dots B(x_{2k},\rho)} \cdots \int (1 - D_r(x_1, \dots, x_{2k}))^{n-2k} \, \mathrm{d}x_1 \, \mathrm{d}x_3 \dots \mathrm{d}x_{2k-1} \\
= \\
\int\limits_{B(x_2,\rho)\dots B(x_{2k},\rho)} \cdots \int (1 - \sum_{i=1}^{k} D_r(x_{2i-1}, x_{2i}) + o(n^{-1}))^{n-2k} \, \mathrm{d}x_1 \dots \mathrm{d}x_{2k} \\
= \\
\int_0^{2\pi} \int_0^{\rho} \dots \int_0^{2\pi} \int_0^{\rho} (1 - \sum_{i=1}^{k} D_r(u_i, \alpha_i) + o(n^{-1}))^{n-2k} u_1 \dots u_k \, \mathrm{d}u_1 \, \mathrm{d}\alpha_1 \dots \mathrm{d}u_k \, \mathrm{d}\alpha_k \\
= \\
(1 + o(1)) \int_0^{2\pi} \int_0^{\rho} \dots \int_0^{2\pi} \int_0^{\rho} \Pi_{i=1}^{k} (1 - D_r(u_i, \alpha_i))^{n-2k} u_1 \dots u_k \, \mathrm{d}u_1 \, \mathrm{d}\alpha_1 \dots \mathrm{d}u_k \, \mathrm{d}\alpha_k \\
= \\
(1 + o(1)) \left(\int_0^{2\pi} \int_0^{\rho} (1 - D_r(u, \alpha))^{n-2k} u \, \mathrm{d}u \, \mathrm{d}\alpha \right)^k \\
= \\
(1 + o(1)) \left(\int_0^{2\pi} \int_0^{\rho} (1 - D_r(u, \alpha))^{n-2} u \, \mathrm{d}u \, \mathrm{d}\alpha \right)^k \\
= \\
(1 + o(1)) \left(\frac{2\mu(r)}{n^2} \right)^k ,
\end{array}$$

where the fourth line follows from the familiar change of variables. The fifth line from the fact that $\Pi_{i=1}^{k}(1 - D_r(u_i, \alpha)) = 1 - \sum_{i=1}^{k} D_r(u_i, \alpha) + o(n^{-1})$, using that $D_r(u_i, \alpha) = \mathcal{O}(\rho^2 r^2) = o(n^{-\frac{1}{2}})$ by Lemma 16, so that $\left(\frac{1 - \sum_{i=1}^{k} D_r(u_i,\alpha) + o(n^{-1})}{\Pi_{i=1}^{k}(1 - D_r(u_i,\alpha))} \right)^n = 1 + o(1)$. The change in exponent in the sixth line follows from $1 - D_r(u, \alpha) = 1 - \mathcal{O}(\rho r) = 1 - o(1)$, and the last line follows from the induction hypothesis. Thus, setting

$$W := \{(x_2, x_4, \dots, x_{2k}) \in ([0, 1]^2)^k : \| x_{2i} - x_{2j} \| > \rho' \text{ for all } 1 \le i < j \le k\} ,$$

we also have

$$\begin{aligned}
\mathbb{P}(A(k)) &= \int \cdots \int_W \mathbb{P}(A(k)|X_2 = x_2, \dots, X_{2k} = x_{2k}) \, \mathrm{d}x_2 \, \mathrm{d}x_4 \dots \mathrm{d}x_{2k} \\
&= (1 + o(1)) \left(\frac{2\mu(r)}{n^2} \right)^k ,
\end{aligned}$$

since the $2k$-dimensional volume $\mathrm{vol}(W)$ of W is $1 + o(1)$ and the error terms in (12) and further are uniform over all $(x, \dots, x_{2k}) \in W$. Combining this with (11) gives the result. □

Here we should mention that the case where $k > 1$ in the proof of Lemma 14 follows the lines of an argument of Agarwal and Spencer [1], but we have included it for completeness.

Proof of Lemma 16. We need to consider the area $D_r(u, \alpha) = \mathrm{vol}(B(x, r) \Delta B(y, r))$ for x, y with $\| x - y \| = u$ and the angle between $y - x$ and the diagonal $\{(a, a)^T : a \in \mathbb{R}\}$ is α.

For ease of computation, let us work with $\left[-\frac{1}{2}, \frac{1}{2}\right]^2$ instead of $[0,1]^2$ in this proof, and we may assume without loss of generality that $x = (0,0)^T$ is the centre of the square. Let v_α be a unit vector that makes an angle α with the diagonal of the unit square, and let w_α be perpendicular to v_α. Let us first consider $0 < r < \frac{1}{2}$. In this case, for u small enough, $B(x,r)\Delta B(y,r)$ lies completely in the interior of the unit square (so there are no effects due to the toroïdal assumption). Let S be the boundary of $B(x,r)$, and set $H_\alpha(c) := \{p : p.w_\alpha = c\}$. So $H_\alpha(c)$ is a line parallel to v_α. We shall approximate $D_r(u,\alpha)$ by $\mathrm{vol}\,(S + [0,u]v_\alpha) \le 4ur$. Note that the height of S is $2r$. Also, observe that for most $c \in (-r,r)$, the set $(B(x,r)\Delta B(y,r)) \cap H_\alpha(c)$ and the set $(S + [0,u]v_\alpha) \cap H_\alpha(c)$ both consist of two line segments, each of length u. It is not hard to see that the c for which this is not the case are contained in $(-r,-r+u) \cup (r-u,r)$, so that

$$4ur \ge D_r(u,\alpha) \ge 4ur - 2u^2\,.$$

This concludes the proof when $0 < r < \frac{1}{2}$.

We now assume that $\frac{1}{2} \le r < \frac{1}{2}\sqrt{2}$. We shall proceed in a similar manner. Again, let S be the boundary of $B(x,r)$. But note that now, due to the toroïdal assumption, S consists of four arcs of opening angle β (see figure 8). We again wish to approximate $D_r(u,\alpha)$ by

Figure 8: The "boundary" of $B(x,r)$ consists of four arcs of opening angle β (left), and the projections of the four arcs onto $\mathcal{L}(w_\alpha)$ (right).

the area of $S + [0,u]v_\alpha$. Let $h(\alpha)$ be the length "counting multiplicities" of the projection of S onto $\mathcal{L}(w_\alpha)$, i.e. $h(\alpha) := \int_{\mathbb{R}} |H_\alpha(c) \cap S|\,\mathrm{d}c$.
We assert that

$$D_r(u,\alpha) = uh(\alpha) + o(u)\,.$$

To see this, note that the length of $(B(x,r)\Delta B(y,r)) \cap H_\alpha(c)$ equals u times the cardinality of $S \cap H_\alpha(c)$, unless one or more of the points in $S \cap H_\alpha$ are

a) within u of the boundary of the square, or

b) within u of another point of $S \cap H_\alpha$.

We have already seen that the error due to b) can be bounded by $2u^2$. In order to bound the error due to a), let $S'(u)$ be the set of all $s \in S$ for which a) is not the case. As u tends to 0 the length $l(S'(u))$ of $S'(u)$ tends to the length $l(S)$ of S. Hence we see that

$$uh(\alpha) - u(l(S) - l(S'(u))) - 2u^2 \leq D_r(u,\alpha) \leq uh(\alpha)\,,$$

as required. Since $l(S) - l(S'(u))$ does not depend on α the error term is indeed uniform in α.

It only remains to compute $h(\alpha)$. For $\alpha \in \left(-\frac{\pi}{4}, -\frac{\beta}{2}\right)$, the length of the projections on $\mathcal{L}(w_\alpha)$ of the two arcs that contain the diagonal of the square is equal to

$$r\left(\sin\left(\frac{\beta}{2} - \alpha\right) - \sin\left(-\frac{\beta}{2} - \alpha\right)\right)\,,$$

while the height of the other two arcs is equal to

$$r\left(\sin\left(\frac{\pi}{2} - \frac{\beta}{2} - \alpha\right) - \sin\left(\frac{\pi}{2} + \frac{\beta}{2} - \alpha\right)\right)\,.$$

Thus, for $-\frac{\pi}{4} < \alpha < -\frac{\beta}{2}$ we obtain

$$\begin{aligned} h(\alpha) &= 2r\left(\sin\left(\tfrac{\beta}{2} - \alpha\right) - \sin\left(-\tfrac{\beta}{2} - \alpha\right) + \sin\left(\tfrac{\pi-\beta}{2} - \alpha\right) - \sin\left(\tfrac{\pi+\beta}{2} - \alpha\right)\right) \\ &= 2r\left(\sin\left(\tfrac{\beta}{2} - \alpha\right) - \sin\left(-\tfrac{\beta}{2} - \alpha\right) + \cos\left(-\tfrac{\beta}{2} - \alpha\right) - \cos\left(\tfrac{\beta}{2} - \alpha\right)\right) \\ &= 4r\sin\left(\tfrac{\beta}{2}\right)\left(\cos(\alpha) - \sin(\alpha)\right)\,, \end{aligned}$$

and for $-\frac{\beta}{2} < \alpha < \frac{\beta}{2}$, we obtain

$$\begin{aligned} h(\alpha) &= 2r\left(\sin\left(\tfrac{\beta}{2} - \alpha\right) - \sin\left(-\tfrac{\beta}{2} - \alpha\right) + 2 - \sin\left(\tfrac{\pi-\beta}{2} - \alpha\right) - \sin\left(\tfrac{\pi+\beta}{2} - \alpha\right)\right) \\ &= 2r\left(\sin\left(\tfrac{\beta}{2} - \alpha\right) - \sin\left(-\tfrac{\beta}{2} - \alpha\right) + 2 - \cos\left(-\tfrac{\beta}{2} - \alpha\right) - \cos\left(\tfrac{\beta}{2} - \alpha\right)\right) \\ &= 4r\left(1 - \left(\cos\left(\tfrac{\beta}{2}\right) - \sin\left(\tfrac{\beta}{2}\right)\right)\cos(\alpha)\right)\,. \end{aligned}$$

This concludes the proof. $\square$

We should mention here that for $r \in \left(0, \frac{1}{2}\right)$, the result can also be obtained in a relatively straightforward manner by explicitly computing $D_r(u,\alpha) = 2\pi r^2 - 4r^2 \arccos\left(\frac{u}{2r}\right) + 2u\sqrt{r^2 - \frac{u^2}{4}}$ and considering the Taylor expansion of this expression. We have not chosen this route because the method used fits better with the case where $r \in \left(\frac{1}{2}, \frac{\sqrt{2}}{2}\right)$.

3.2 Locating-dominating codes

We prove the following result.

Theorem 17. *Let M be the cardinality of a smallest locating-dominating code.*

(i) If $nr^2 \ll \ln n$ then $M = n^{1+o(1)}$ with high probability;

(ii) if $n^{-\frac{1}{2}} \ll r \le \frac{1}{2}\sqrt{2} - \varepsilon$ *for some* $\varepsilon > 0$ *then* $M = \left(\frac{n}{r}\right)^{\frac{2}{3}+o(1)}$ *with high probability.*

Our computations below use the following lemma, which is a reformulation by Janson [10] of an inequality due to Suen [17].

Lemma 18. *Let* $A_1, \dots, A_m$ *be events, and let* H *be a graph with vertex set* $\{1, \dots, m\}$ *and an edge* $ij \in E(H)$ *whenever* A_i *and* A_j *are dependent. Let* $Z := \sum_{i=1}^m 1_{A_i}$ *be the number of events that hold and set* $\mu := \mathbb{E}Z, \Delta := \sum_{ij\in E(H)} \mathbb{P}(A_i \cap A_j), \delta := \max_i \sum_{j:ij\in E(H)} \mathbb{P}(A_j)$*. Then it holds that*

$$\mathbb{P}(Z = 0) \le \exp[-\mu + \Delta e^{2\delta}]\,.$$

Proof of Theorem 17. Part (i). The statement follows by showing that whp. there are $n^{1+o(1)}$ isolated vertices, because a locating-dominating code must contain all the isolated vertices. Let Y be the number of isolated vertices. Then

$$\mathbb{E}Y := n(1-\pi r^2)^{n-1} = ne^{(n-1)\ln(1-\pi r^2)} = ne^{-(\pi r^2+\mathcal{O}(r^4))(n-1)} = ne^{o(\ln n)} = n^{1+o(1)}\,.$$

Let us now compute $\mathrm{Var}(Y)$. Note that $\mathrm{Var}(Y)$ is equal to

$$\begin{aligned}&\sum_i\sum_j \left(\mathbb{P}(X_i, X_j \text{ both isolated}) - \mathbb{P}(X_i \text{ isolated})\mathbb{P}(X_j \text{ isolated})\right)\\ =&\mathbb{E}Y(1-\mathbb{P}(X_1 \text{ isolated})) + n(n-1)\left(\mathbb{P}(X_1, X_2 \text{ both isolated}) - \mathbb{P}(X_1 \text{ isolated})^2\right)\,.\end{aligned}$$

Also, notice that

$$\begin{aligned}\mathbb{P}(X_1, X_2 \text{ both isolated }) &\le (1-\pi r^2)^{n-1}4\pi r^2 + (1-2\pi r^2)^{n-2}\\ &\le (1-\pi r^2)^{n-1}4\pi r^2 + \tfrac{1}{(1-\pi r^2)^2}(1-\pi r^2)^{2(n-1)}\\ &= 4\pi r^2\mathbb{P}(X_1 \text{ isolated}) + \left(1 - \tfrac{\pi^2 r^4 - 2\pi r^2}{(1-\pi r^2)^2}\right)\mathbb{P}(X_1 \text{ isolated})^2\,.\end{aligned}$$

Thus, $\mathbb{P}(X_1, X_2 \text{ both isolated}) - \mathbb{P}(X_1 \text{ isolated})^2 = \mathcal{O}(r^2\mathbb{P}(X_1 \text{ isolated}))$. As $r^2 = o\left(\frac{\ln n}{n}\right)$, we thus find that

$$\mathrm{Var}(Y) = o(\ln n \cdot \mathbb{E}Y) = o((\mathbb{E}Y)^2)\,.$$

So Chebyschev's inequality gives that indeed $Y = (1+o(1))\mathbb{E}Y = n^{1+o(1)}$ whp.

Part (ii). Let $\varepsilon > 0$ be arbitrary. First, we shall construct a code C of cardinality $\mathcal{O}(n^{\frac{2}{3}+\varepsilon})$ as follows. We start with $C_0 := \{X_1, \dots, X_c\}$ with $c = \lceil(\frac{n}{r})^{\frac{2}{3}+\varepsilon}\rceil$, and then we obtain C by adding to C_0 all points in $X_{c+1}, \dots, X_n$ that are in a bad pair for C_0, i.e., if $C_0 \cap \overline{N}(X_i) = C_0 \cap \overline{N}(X_j)$ for $c < i < j \le n$ then we add X_i and X_j. Let Y be the number of pairs in $X_{c+1}, \dots, X_n$ that are bad with respect to C. Notice that now a bad pair need not correspond to an edge of G_n. We see that

$$\begin{aligned}\mathbb{E}Y &= \tbinom{n-c}{2}\mathbb{P}\left(\{X_{c+1}, X_{c+2}\} \text{ bad for } C_0\right)\\ &= (1+o(1))\tfrac{n^2}{2}\left(\int_0^{2r}\int_0^{2\pi}(1-D_r(u,\alpha))^c u\,\mathrm{d}\alpha\,\mathrm{d}u + \mathcal{O}\left(e^{-c2\pi r^2}\right)\right)\\ &= (1+o(1))\tfrac{n^2}{32c^2r^2}\left(\int_0^{2\pi} F(\alpha)^{-2}\,\mathrm{d}\alpha\right) + \mathcal{O}\left(n^2e^{-c2\pi r^2}\right)\\ &= \Theta\left(\left(\tfrac{n}{r}\right)^{\frac{2}{3}-2\varepsilon}\right)\,.\end{aligned}$$

To obtain the third line we have reused the computations that gave (5), (6), (7) and (8). To obtain the fourth line we have used that $e^{-c2\pi r^2} = o(c^{-2}r^{-2})$, because $c^{-2}r^{-2} = (1+o(1))(\frac{n}{r})^{-2\varepsilon}/n(nr^2)^{\frac{1}{3}} \gg n^{-100}$ and $cr^2 = (1+o(1))(nr^2)^{\frac{2}{3}}(\frac{n}{r})^{\varepsilon} \gg n^{\varepsilon}$, so that $e^{-c2\pi r^2} = o(e^{-n^{\varepsilon}})$.

By Markov's inequality,

$$\mathbb{P}(Y \geq n^{\varepsilon} \cdot \mathbb{E}Y) \leq n^{-\varepsilon}\,.$$

Thus, $Y = \mathcal{O}((\frac{n}{r})^{\frac{2}{3}-\varepsilon})$ whp. This gives that there indeed is a code of size $|C| = c + Y = \mathcal{O}((\frac{n}{r})^{\frac{2}{3}+\varepsilon})$ whp.

Now let us consider a lower bound. We set $c := \lfloor (\frac{n}{r})^{\frac{2}{3}-\varepsilon} \rfloor$. Observe that

$$\mathbb{P}(M \leq c) \leq \binom{n}{c} \mathbb{P}(X_1, \ldots, X_c \text{ is a locating-dominating code})\,.$$

As before, we set $\mathcal{P} := \binom{\{X_{c+1},\ldots,X_n\}}{2}$ and $P_{ij} := \{X_{c+i}, X_{c+j}\}$. We call a pair $P_{ij} \in \mathcal{P}$ a *close bad pair* if P_{ij} is bad and $\| X_{c+i} - X_{c+j} \| < \rho := \frac{1}{cr}$. Let Z be the number of close bad pairs. Then, the probability that $\{X_1, \ldots, X_c\}$ is a locating-dominating code is at most the probability that Z is zero, which in turn is at most

$$\sup_{x_1,\ldots,x_c \in [0,1]^2} \mathbb{P}(Z = 0 | X_1 = x_1, \ldots, X_c = x_c)\,.$$

We use Lemma 18 above. Let us fix $x_1, \ldots, x_c \in [0,1]^2$ and let the random variable $\tilde{Z}$ satisfy $\tilde{Z} \stackrel{d}{=} (Z | X_1 = x_1, \ldots, X_c = x_c)$. Notice that if we condition on the vector $(X_i - X_j)$,then the probability that $X_k \in B(X_i, r)\Delta B(X_j, r)$ is exactly $D_r(u, \alpha)$ (with $u = \| X_i - X_j \|$ and α the angle of $X_i - X_j$ with the diagonal of the unit square). Also notice that $D_r(u,\alpha) \leq 4ru$ uniformly in r, u and α. Hence,

$$\begin{aligned}
\mu &:= \mathbb{E}\tilde{Z} \geq \tbinom{n-c}{2} \int_0^{2\pi} \int_0^{\rho} (1 - cD_r(u,\alpha)) u \,\mathrm{d}u \,\mathrm{d}\alpha, \\
&\geq (1+o(1)) \tfrac{n^2}{2} \left[\pi u^2 - \tfrac{8\pi}{3} cru^3\right]_0^{\rho} \\
&= (1+o(1)) \tfrac{\pi n^2 \rho^2}{2} \left(1 - \mathcal{O}\left(rc\rho^3\right)\right) \\
&= \Omega\left(\left(\tfrac{n}{cr}\right)^2\right) = \Omega\left(\left(\tfrac{n}{r}\right)^{\frac{2}{3}+2\varepsilon}\right) = \Omega\left(cn^{3\varepsilon}\right)\,.
\end{aligned}$$

Now let us consider

$$\begin{aligned}
\Delta &:= \sum_{\substack{P,P' \in \mathcal{P} \\ P_1 \cap P_2 \neq \emptyset}} \mathbb{P}(P, P' \text{ both close bad} | X_1 = x_1, \ldots, X_c = x_c) \\
&= \tbinom{n-c}{2} 2(n-c-2) \mathbb{P}(P_{12}, P_{13} \text{ close bad} | X_1 = x_1, \ldots, X_c = x_c) \\
&\leq \mu\left(\pi n \rho^2\right) \leq \mu\left(\pi n c^{-2} r^{-2}\right) = \mathcal{O}\left(\mu \left(nr^2\right)^{-\frac{1}{3}+\varepsilon} r^{\varepsilon}\right) = o\left(\mu\right)\,.
\end{aligned}$$

Let us set

$$\delta := \max_{P \in \mathcal{P}} \sum_{\substack{P' \in \mathcal{P}, \\ P \cap P' \neq \emptyset}} \mathbb{P}(P, P' \text{ both close bad} | X_1 = x_1, \ldots, X_c = x_c)\,.$$

By symmetry considerations, we obtain

$$\delta \;=\; \frac{\Delta}{\binom{n-c}{2}} = o(\tfrac{\mu}{n^2}) = o(1)\,.$$

Thus, by Lemma 18, we find

$$\mathbb{P}(Z=0|X_1=x_1,\ldots,X_n=x_n) = \mathbb{P}(\tilde{Z}=0) \le \exp[-\mu+\Delta e^{\delta}] = \exp[-\Omega(n^{\frac{2}{3}+2\varepsilon})]\,.$$

Note also that this bound is uniform in $x_1,\ldots,x_n$ so that the right hand side is in fact also an upper bound for $\mathbb{P}(\{X_1,\ldots,X_c\}$ is a code). We see that

$$\begin{aligned}\mathbb{P}(M\le c) &\le \tbinom{n}{c}\mathbb{P}(X_1,\ldots,X_c \text{ is a locating-dominating code})\\ &\le n^c\exp[-\Omega(cn^{3\varepsilon})]\\ &= \exp[\mathcal{O}(c\ln n)-\Omega(cn^{3\varepsilon})] = o(1)\,.\end{aligned}$$

This concludes the proof. □

4 Further work

In this paper we have always assumed that the vertices of the random unit disk graph follow the uniform distribution on the unit square, which we made into a torus by identifying opposite sides. A possibly non-trivial exercise would be to extend the proofs to other probability distributions on the plane, and to derive the exact limiting probabilities of the existence of an identifying code for fixed r without the toroïdal assumption

Another topic for future research, especially motivated by the applications, is to investigate approximation algorithms for finding minimum identifying codes or locating-dominating codes for unit disk graphs.

5 Acknowledgement

We would like to thank the referee for careful proofreading and many helpful comments which have greatly improved the paper.

References

[1] A. Agarwal and J. Spencer. Undecidable statements and the zero-one law in random geometric graphs. Preprint, 2006.

[2] H. Breu and D. G. Kirkpatrick. Unit disk graph recognition is NP-hard. *Comput. Geom.*, 9(1-2):3–24, 1998.

[3] K. Chakrabarty, M. G. Karpovsky, and L. B. Levitin. Fault isolation and diagnosis in multiprocessor systems with point-to-point connections. In D. R. Avresky and D. R. Kaeli, editors, *Fault-Tolerant Parallel and Distributed Systems*, pages 285–301. Kluwer Academic Publishers, 1998.

[4] I. Charon, O. Hudry, and A. Lobstein. Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard. *Theoret. Comput. Sci.*, 290(3):2109–2120, 2003.

[5] B. N. Clark, C. J. Colbourn, and D. S. Johnson. Unit disk graphs. *Discrete Math.*, 86(1-3):165–177, 1990.

[6] C. J. Colbourn, P. J. Slater, and L. K. Stewart. Locating dominating sets in series parallel networks. *Congr. Numer.*, 56:135–162, 1987. Sixteenth Manitoba conference on numerical mathematics and computing (Winnipeg, Man., 1986).

[7] U. Fößmeier, G. Kant, and M. Kaufmann. 2-visibility drawings of plane graphs. In *Proceedings of Graph Drawing '96*, volume 1190 of *Lecture Notes in Comput. Sci.*, pages 155–168. Springer, Berlin, 1996.

[8] A. Frieze, R. Martin, J. Moncel, M. Ruszinko, and C. Smyth. Codes identifying sets of vertices in random networks. *Discrete Math.*, 307(9–10):1094–1107, 2007.

[9] I. Honkala, M. G. Karpovsky, and L. B. Levitin. On robust and dynamic identifying codes. *IEEE Trans. Inform. Theory*, 52(2):599–612, 2006.

[10] S. Janson. New versions of Suen's correlation inequality. In *Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997)*, volume 13, pages 467–483, 1998.

[11] M. G. Karpovsky, K. Chakrabarty, and L. B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Trans. Inform. Theory*, 44(2):599–611, 1998.

[12] D. Lichtenstein. Planar formulae and their uses. *SIAM J. Comput.*, 11(2):329–343, 1982.

[13] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson. Wireless sensor networks for habitat monitoring. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 88–97, 2002.

[14] A. Papakostas and I. G. Tollis. Orthogonal drawings of high degree graphs with small area and few bends. In *Proceedings of 5th Workshop on Algorithms and Data Structures*, volume 1272 of *Lecture Notes in Comput. Sci.*, pages 354–367. Springer, Berlin, 1998.

[15] M. D. Penrose. *Random Geometric Graphs*. Oxford University Press, Oxford, 2003.

[16] S. Ray, D. Starobinski, A. Trachtenberg, and R. Ungrangsi. Robust location detection with sensor networks. *IEEE Journal on Selected Areas in Communications (Special Issue on Fundamental Performance Limits of Wireless Sensor Networks)*, 22(6):1016–1025, 2004.

[17] W.-C. S. Suen. A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Structures Algorithms*, 1(2):231–242, 1990.